BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Software

Volume 10(7)

MICO: A meta-tool for prediction of the effects of non-synonymous mutations

Gilliean Lee[1] & Chin-Fu Chen[2]*

[1]Department of Mathematics & Computing, Lander University, Greenwood, SC, 29649; [2]Center for Molecular Studies and Office of Bioinformatics and Epidemiology, Greenwood Genetic Center, Greenwood, SC, 29646; Chin-Fu Chen – Email: cfchen@ggc.org; *Corresponding author

Received June 20, 2014; Accepted June 27, 2014; Published July 22, 2014

Abstract:
The Next Generation Sequencing (NGS) is a state-of-the-art technology that produces high throughput data with high resolution mutation information in the genome. Numerous methods with different efficiencies have been developed to predict mutational effects in the genome. The challenge is to present the results in a balanced manner for better biological insights and interpretation. Hence, we describe a meta-tool named Mutation Information Collector (MICO) for automatically querying and collecting related information from multiple biology/bioinformatics enabled web servers with prediction capabilities. The predicted mutational results for the proteins of interest are returned and presented as an easy-to-read summary table in this service. MICO also allows for navigating the result from each website for further analysis.

Availability: http: //mico.ggc.org /MICO

Background:
NGS technology can generate nearly complete information of genetic mutations in the human genome. However, the enormous amount of information also represents a great challenge for a researcher in comprehending the significance of all these mutations. For example, the estimated number of non-synonymous single-nucleotide variants (SNVs) in each human is 24,000- 40,000 **[1]**, which is almost impossible for a regular laboratory to sort through using experimental approaches.

A useful approach for solving the problem of too many SNVs for interpretation is to utilize computational methods and predict *in silico* the consequences of these point mutations. Many bioinformatic tools have been developed in the past few years for this purpose **[2-4]**. However, different algorithms could generate different predictions regarding the consequence of the mutation on a given protein. We believe that a better strategy is to gather information from as many different computational tools as possible. We developed MICO web interface that contains six leading prediction tools: **Condel**, **MutationAssesor, Mutation Taster**, **PolyPhen2**, **SIFT,** and **CADD** see **Figure 1**.

Condel uses a weighted average of the normalized scores (WAS) for integration of five prediction tools -*Logre*, *MAPP*, *MutationAssessor*, *Polyphen2*, and *SIFT*- into a unified classification **[5]**. **MutationAssesor** calculates the change of entropy of a refined class of evolutionarily conserved residues with functional specificity to predict the mutation impact **[6]**. **MutationTaster** implements a naïve Bayes classifier for evaluation of the mutation potential using the information obtained from multiple biomedical databases and prediction methods such as *NNSplice*, *polyadq*, *Grantham Matrix*, *phastCons* and *phyloP* **[7]**. **PolyPhen2** determines if a given mutation is damaging by a naïve Bayes posterior probability based on the results of eight sequence-based and three structure-based predictive features **[8]**. **SIFT** utilizes a sequence homology-based algorithm ("sorting tolerant from intolerant") to evaluate amino acid substitutions within protein families **[9]**. **CADD** reports each variant's deleteriousness based on allelic diversity,

functionality, pathogenicity, disease severity, regulatory effects, and complex trait associations **[10].**

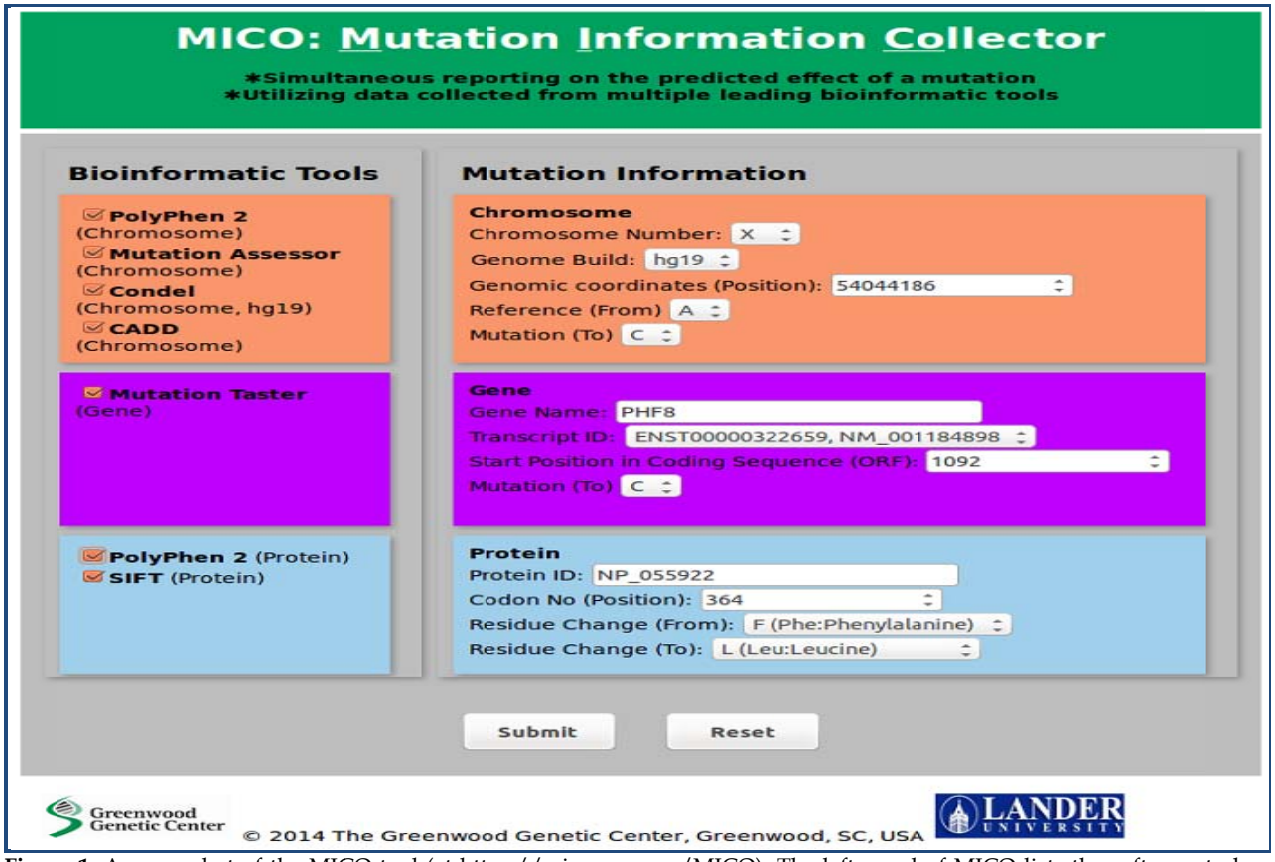
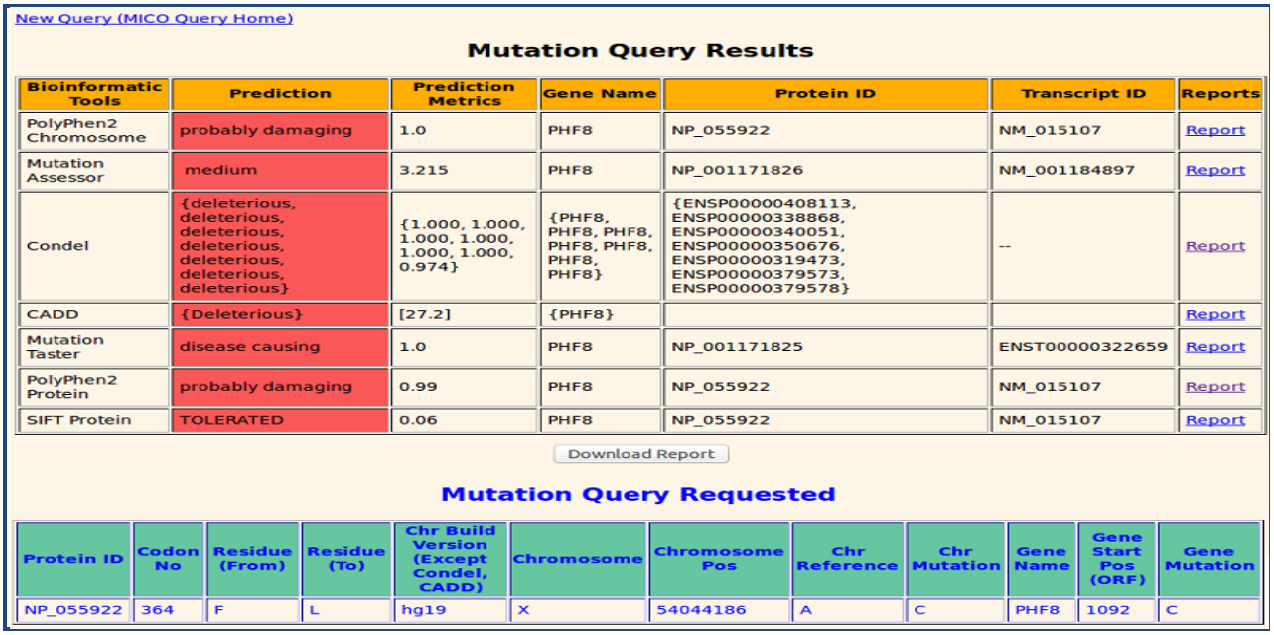


Figure 1: A screenshot of the MICO tool (at http: //mico.ggc.org /MICO). The left panel of MICO lists the software to be used; right panel is for entering the mutation information.



Bioinformatic Tools	Prediction	Prediction Metrics	Gene Name	Protein ID	Transcript ID	Reports
PolyPhen2 Chromosome	probably damaging	1.0	PHF8	NP_055922	NM_015107	Report
Mutation Assessor	medium	3.215	PHF8	NP_001171826	NM_001184897	Report
Condel	{deleterious, deleterious, deleterious, deleterious, deleterious, deleterious, deleterious}	{1.000, 1.000, 1.000, 1.000, 1.000, 1.000, 0.974}	{PHF8, PHF8, PHF8, PHF8, PHF8, PHF8, PHF8}	{ENSP00000408113, ENSP00000338868, ENSP00000340051, ENSP00000350676, ENSP00000319473, ENSP00000379573, ENSP00000379578}	--	Report
CADD	{Deleterious}	[27.2]	{PHF8}			Report
Mutation Taster	disease causing	1.0	PHF8	NP_001171825	ENST00000322659	Report
PolyPhen2 Protein	probably damaging	0.99	PHF8	NP_055922	NM_015107	Report
SIFT Protein	TOLERATED	0.06	PHF8	NP_055922	NM_015107	Report

Protein ID	Codon No	Residue (From)	Residue (To)	Chr Build Version (Except Condel, CADD)	Chromosome	Chromosome Pos	Chr Reference	Chr Mutation	Gene Name	Gene Start Pos (ORF)	Gene Mutation
NP_055922	364	F	L	hg19	X	54044186	A	C	PHF8	1092	C

Figure 2: An example screenshot of a MICO report. Detailed reports from the original bioinformatic tools are available through the Report links in the last column.

Implementation:
MICO submits queries to the bioinformatic tool servers by sending out HTTP requests via the Internet. The status and results of the queries are recognized by parsing the returned HTML documents. MICO consists of a user interface layer, a front-end server layer, a back-end server layer, and a relational database. The front-end server layer includes Web Server/Web Application Server (Apache Tomcat), a Database Management System (MySQL), and MICO Web Application. The Web application is implemented as Java servlets for Web user interface, and a Java library that communicates with back-end servers. The back-end servers are independent bioinformatic tool Web servers for mutation queries. MICO utilizes AJAX (Asynchronous JavaScript And XML) for automatic refresh of query results. The database is used to find matching proteins, transcripts, and chromosomes to supplement query parameters.

Software input:
To minimize the error during data entering process, the query page of MICO first matches the input format with the color-coded bioinformatic tools for their input requirement. After selecting bioinformatic tools, input fields will be automatically enabled. Users can select from predefined values in some fields such as chromosome number, reference, and mutation. MICO automatically checks if the input values are consistent with the required format of the specific web sites.

Software output:
The predicted results from multiple bioinformatic websites are returned and organized as an easy-to-read summary table - "MICO Query Results" (**Figure 2**). The response time from different websites can vary significantly, ranging from a few seconds to several minutes. MICO will update the report every 10 seconds as results become available. If it takes longer than 10 minutes, it wills time out.

The user can now quickly obtain a consensus whether a given mutation may have a deleterious result. The user can also inspect detailed reports originated from the original bioinformatic tools by clicking the underlined "Report" link. The entire search results of MICO can be downloaded as a comma-separated values (csv) file by clicking the "Download Report" button from the result page.

Caveat and future development:
In the near future, we will expand the functionality of MICO to include a total of 37 leading bioinformatic tools. Among these tools, 31 bioinformatic tools will be directly accessed via the Internet as back-end servers, while 6 tools will be implemented locally on our MICO server. Currently, MICO can only accept one mutation per query. We plan to expand the capacity of MICO to accept multiple mutations in batch mode.

Conclusion:
MICO presents the researcher with an unbiased view of all possible predictions on the effects of a given mutation. MICO thus could speed up the understanding of the genetic basis of human diseases. Further, MICO may enhance research in computational biology and bioinformatics. The summary table of MICO would be an easy follow-up for constructing a computational model or algorithm, such as by assigning different weights to each prediction result.

References:

[1] Ng PC & Henikoff S, *Annu Rev Genomics Hum Genet.* 2006 **7:** 61 [PMID: 16824020]
[2] Frousios K *et al. Genomics* 2013 **102:** 223 [PMID: 23831115]
[3] Cline MS & Karchin R, *Bioinformatics* 2011 **27:** 441 [PMID: 21159622]
[4] Thusberg J *et al. Human mutation* 2011 **32:** 358 [PMID: 21412949]
[5] Gonzalez-Perez A *et al. Am J Hum Genet* 2011 **88:** 440 [PMID: 21457909]
[6] Reva B *et al. Nucleic Acids Res* 2011 **39:** e118 [PMID: 21727090]
[7] Schwarz JM *et al. Nat Methods*. 2010 **7:** 575 [PMID: 20676075]
[8] Adzhubei IA *et al. Nat Methods.* 2010 **7:** 248 [PMID: 20354512]
[9] Kumar P *et al. Nat Protoc.* 2009 **4:** 1073 [PMID: 19561590]
[10] Kircher M *et al. Nat Genet.* 2014 **46:** 310 [PMID: 24487276]

Edited by P Kangueane
Citation: **Lee & Chen,** Bioinformation 10(7): 469-471 (2014)